

COMPUTAÇÃO EM NUVEM

Conceitos, tecnologias e aplicações

Jaqueline Joice Brito
Lucas de Carvalho Scabora

Sumário

- O que é Computação em Nuvem?
- Modelos de Serviços
- Prós e Contras
- Aplicações
- BigData
- Bancos de Dados NoSQL
- Hadoop

O que é Computação em Nuvem?

- Computação em nuvem é um modelo que possibilita acesso a um conjunto de recursos computacionais compartilhados e interligados via rede



O que é Computação em Nuvem?

A nuvem é uma **metáfora** para a Internet ou infraestrutura de comunicação entre os componentes arquiteturais

Abstração que oculta a complexidade de infraestrutura



O que é Computação em Nuvem?

Em resumo, é uma plataforma que provê serviços sob-demanda, que está **sempre disponível**, em **qualquer lugar e a qualquer hora**



O que é Computação em Nuvem?

Definição segundo NIST

“Computação em nuvem é um modelo para permitir **acesso ubíquo, conveniente e sob demanda** via rede a um **agrupamento compartilhado e configurável de recursos computacionais** (por exemplo, redes, servidores, equipamentos de armazenamento, aplicações e serviços), **que pode ser rapidamente fornecido e liberado** com esforços mínimos de gerenciamento ou interação com o provedor de serviços.”

O que é Computação em Nuvem?

Principais Características

- Serviço sob demanda
 - Alocação dinâmica de recursos
- Acesso via rede
- Compartilhamento de recursos
- Elasticidade
 - Sensação de capacidade infinita de recursos
- Serviço mensurável
 - Modelo pay-as-you-go

O que é Computação em Nuvem?

	Tecnologia	Modelo de Negócio
Mainframe	Computação Centralizada	Alto custo de hardware e software
Cliente/Servidor	Computação Distribuída	Licença para SO e aplicativos
Computação em Nuvem	Grandes data centers	Custo proporcional ao uso

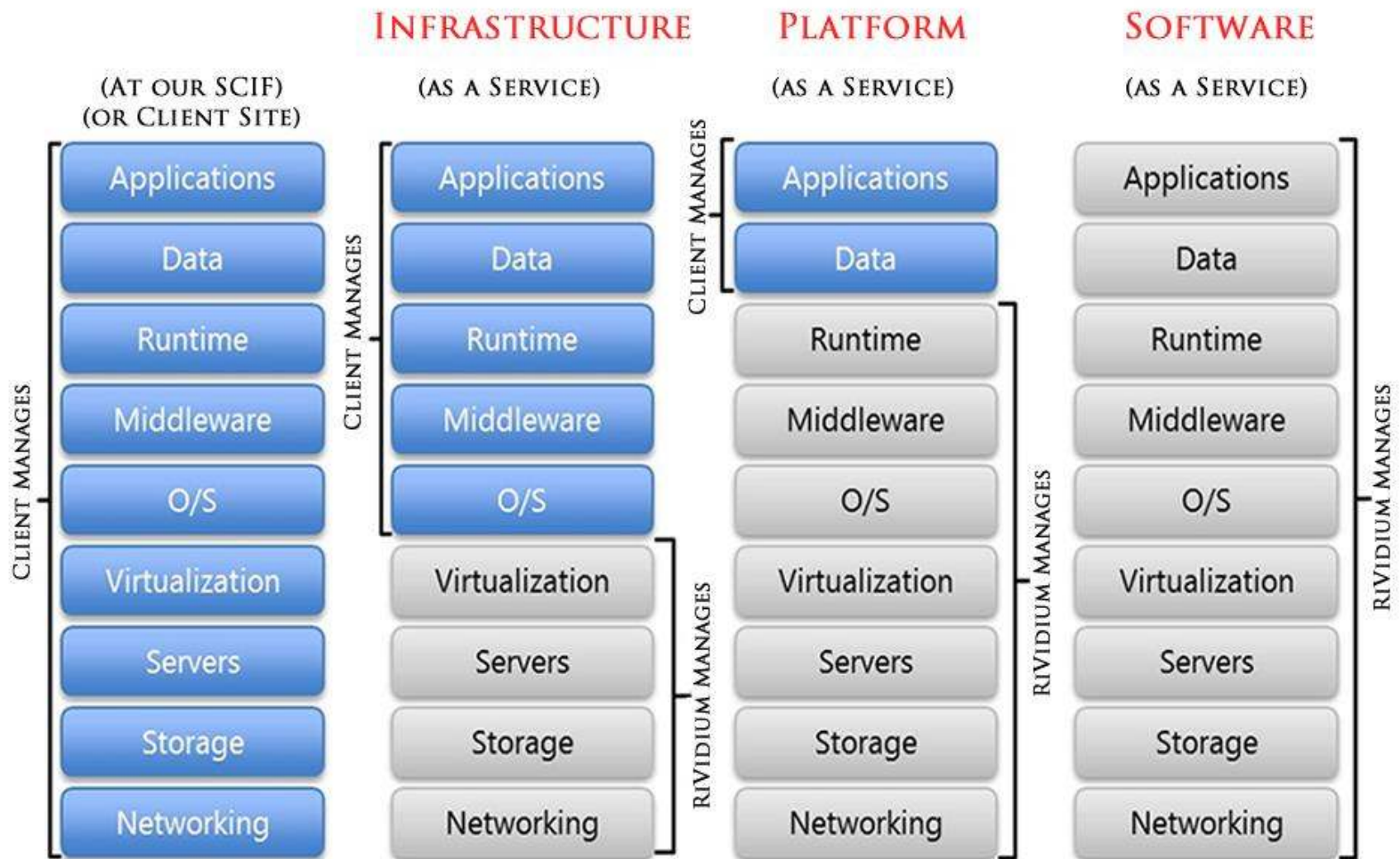
O que é Computação em Nuvem?

Modelos de implantação

- Privada
- Comunitária
- Pública
- Híbrida



Modelos de Serviços



Modelos de Serviços

Termos Emergentes

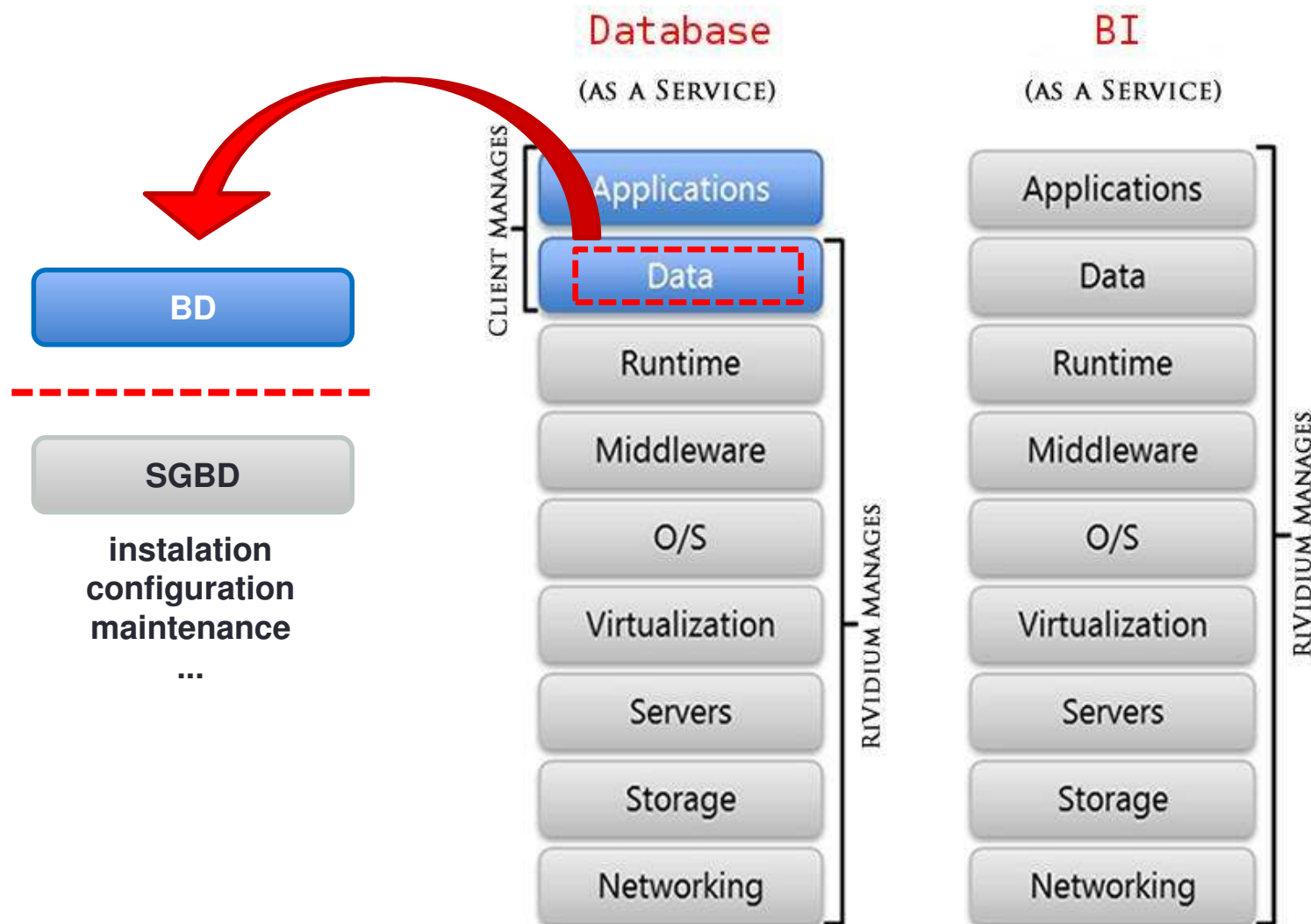
Database as a Service (DBaaS)

O provedor de serviço tem a responsabilidade de instalar e dar manutenção ao banco de dados. Assim, o usuário, que contratou o serviço, apenas paga um valor proporcional ao uso.

Business Intelligence as a Service (BlaaS)







Também chamado de Software como um serviço de business intelligence (SaaS BI), consiste em instalar e gerenciar aplicações de BI na nuvem.

Modelos de Serviços



Modelos de Serviços

Exemplos

Software as a Service (SaaS)	Platform as a Service (PaaS)	Infrastructure as a Service (IaaS)
		
		

Prós e Contras

Principais Vantagens

Redução de investimento em TI (hardware e software)

- Envolvendo custo de manutenção, de pessoal, de espaço físico e de energia

Confiabilidade

- Replicação dos dados
- Disponibilidade das aplicações

Prós e Contras

Principais Vantagens

- Não há contratos complexos e duradouros de prestação de serviço
 - Cobra-se apenas pelos recursos multiplicados pelo tempo de uso
- Escala por demanda
 - Capacidade virtualmente infinitos
- Abstração da tecnologia sendo utilizada

Prós e Contras

Principais Desafios

Segurança e confidencialidade dos dados

Gerenciamento dos dados

Disponibilidade

Integração de serviços

Necessidade de uma banda maior de internet

Aplicações

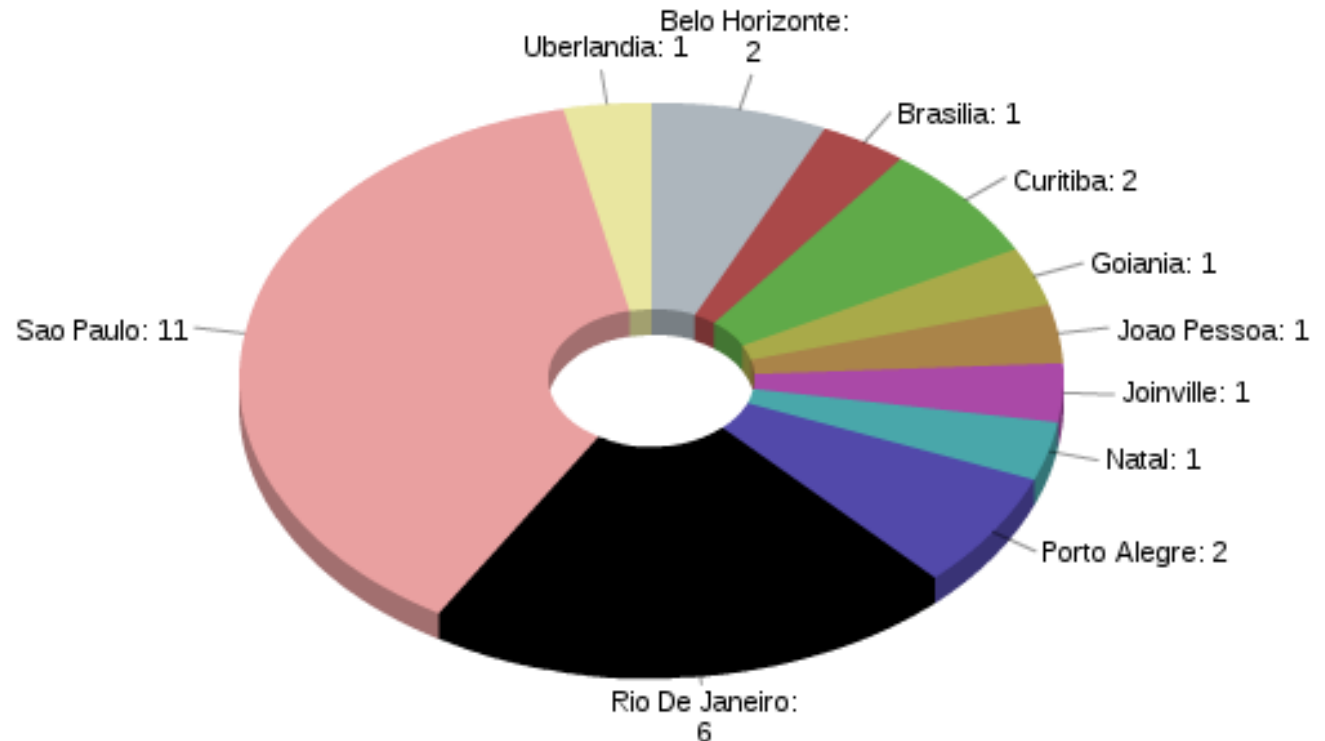
Data Centers



Aplicações

Data Centers

No Brasil, temos ao todo **29 data centers**



Aplicações

ARMAZENAMENTO
COM REDUNDÂNCIA
LOCAL (LRS)



ARMAZENAMENTO
COM REDUNDÂNCIA DE
ZONA (ZRS)



ARMAZENAMENTO
COM REDUNDÂNCIA
GEOGRÁFICA (GRS)



e
ARMAZENAMENTO COM
REDUNDÂNCIA GEOGRÁFICA NO
ACESSO DE LEITURA (RA-GRS)

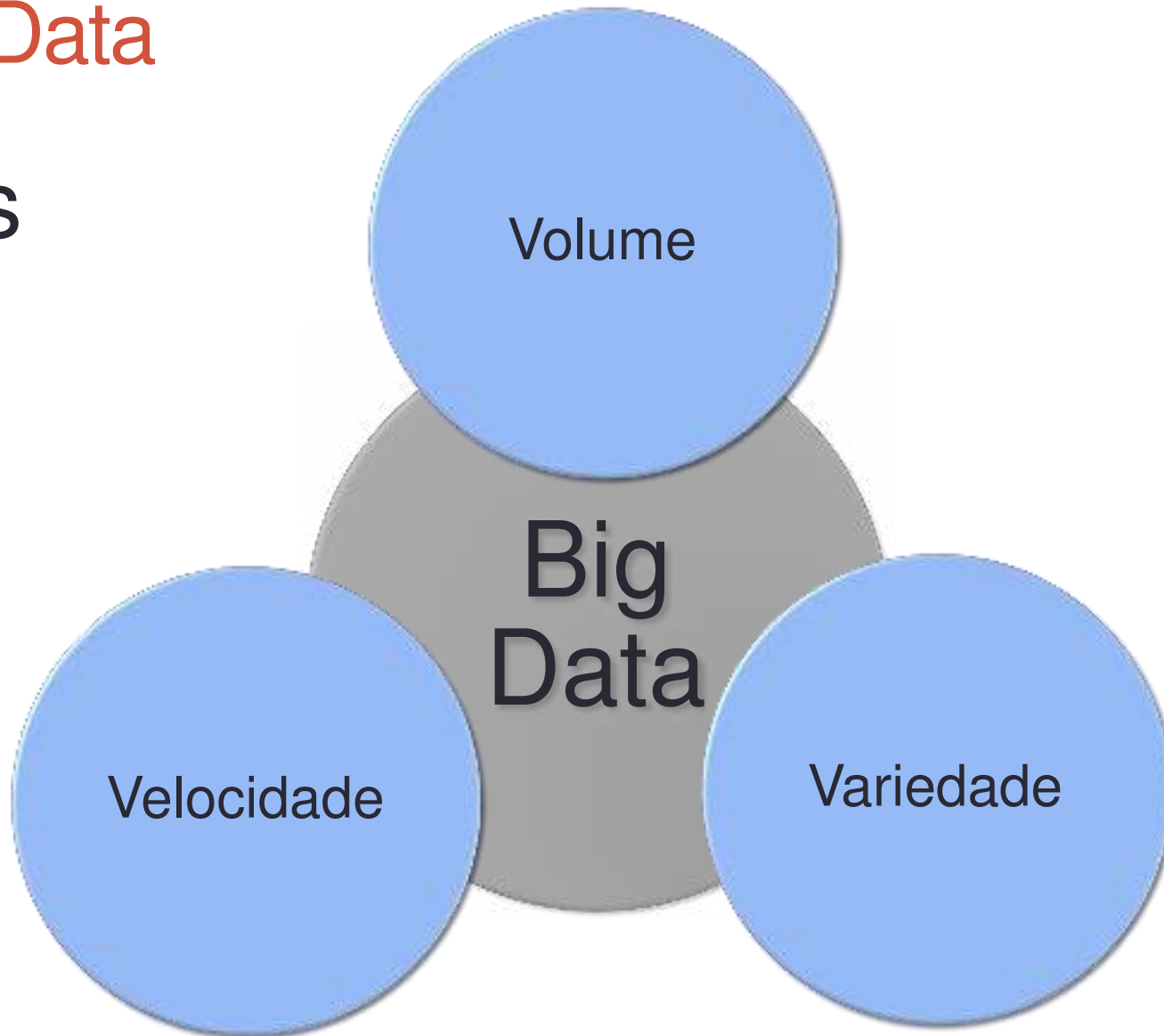
Aplicações

Usos mais comuns da computação em nuvem

- › IaaS e PaaS
- › Armazenamento de dados
- › Ambientes de testes e desenvolvimento
- › **Processamento e análise de Big Data**
- › Backup

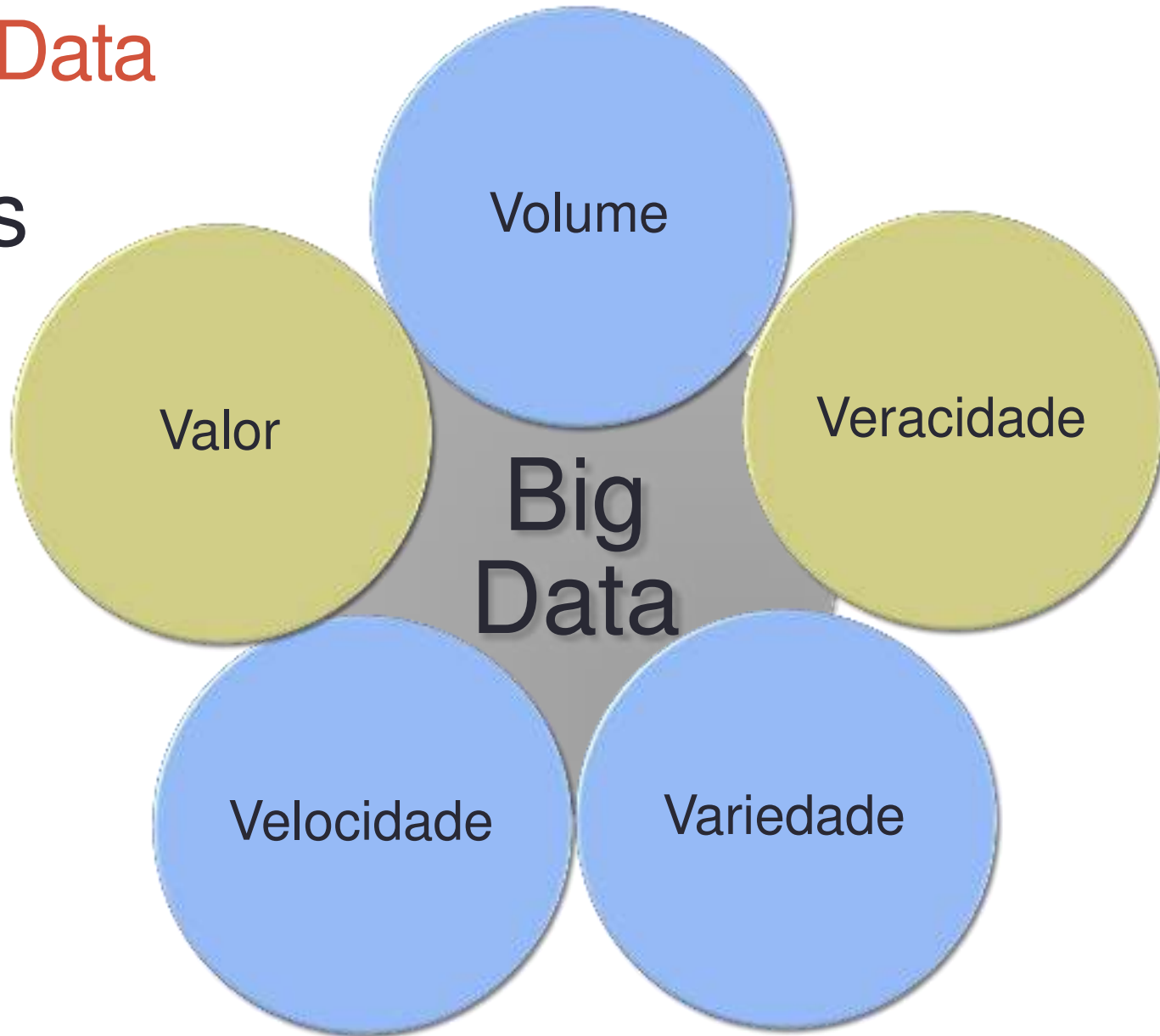
Big Data

3Vs



Big Data

5Vs



Big Data

Dados possuem informações valiosas

- Análise de navegação e reviews de usuários
 - Recomendação personalizada de produtos
 - Análise de mensagens de redes sociais



Big Data

Observatório da Dengue

- Desenvolvido por pesquisadores da UFMG
- Monitoramento de mensagens em redes sociais
- Aumento de mensagens relacionadas ao termo “dengue” em uma região pode indicar um possível surto

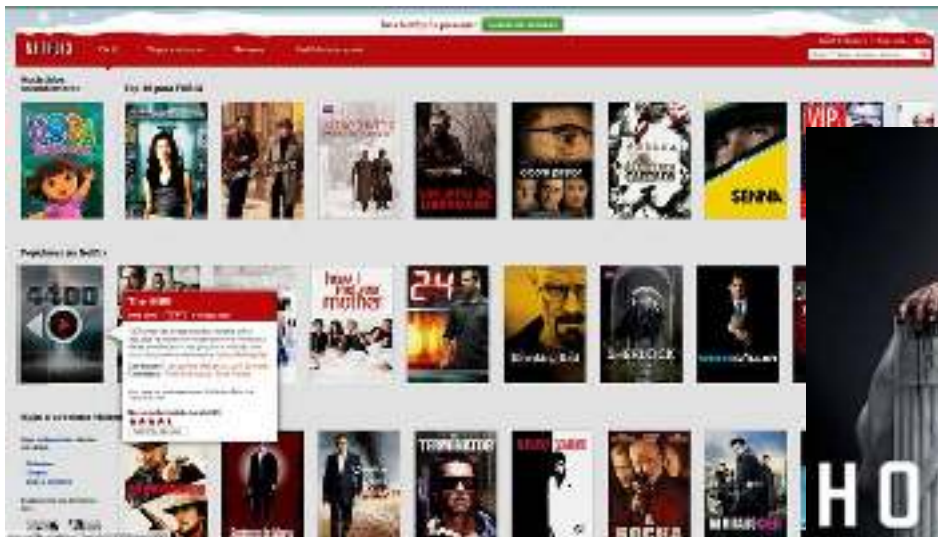


Big Data



Kevin Spacey e David Fincher possuíam grande aceitação

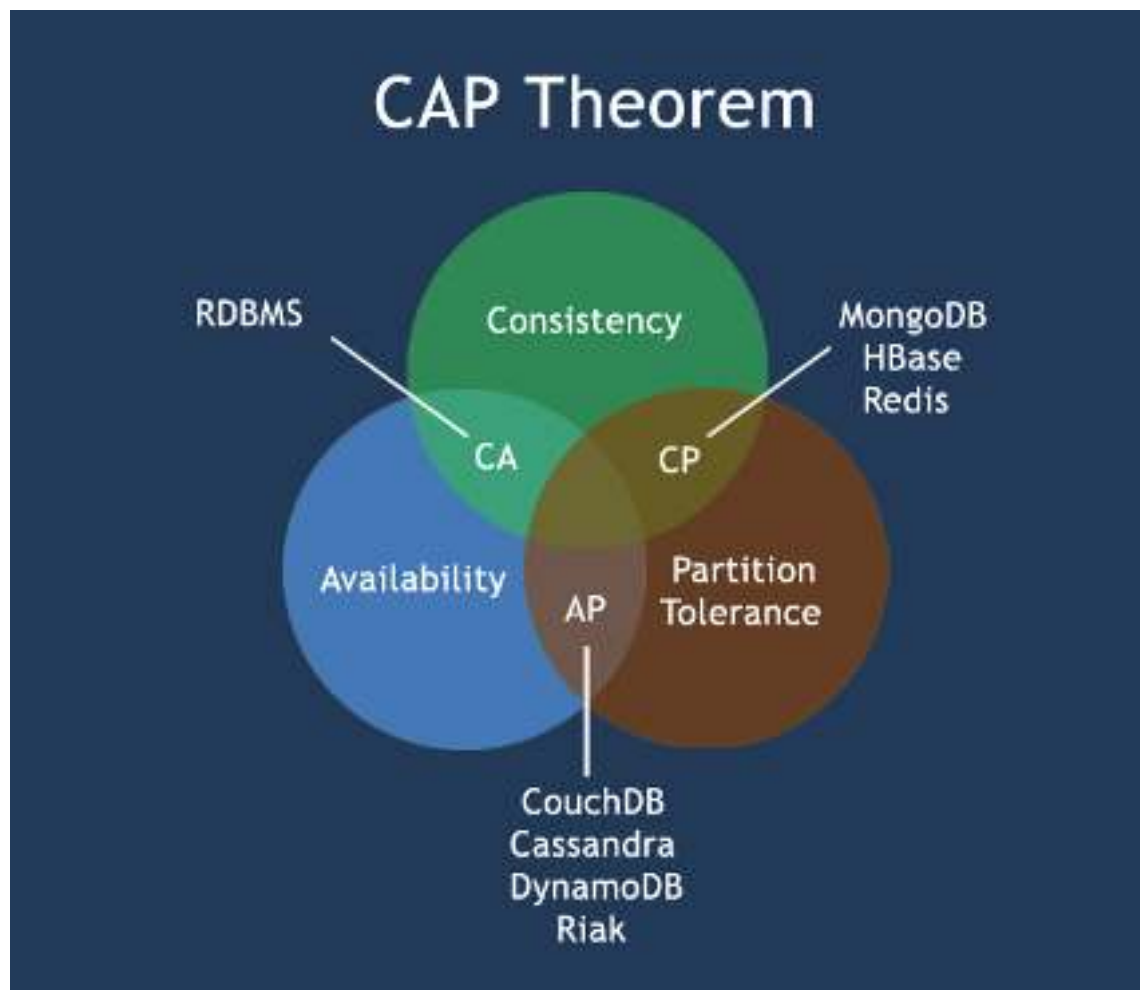
Mesmos usuários gostavam de thrillers políticos



Armazenamento Distribuído de Dados

- Bancos de dados relacionais
 - Dados estruturados
 - Escopo limitado
 - Propriedades ACID
- Bancos de dados NoSQL (Not only SQL)
 - Dados semiestruturados e não estruturados
 - Esquema flexível ou nenhum esquema
 - Relaxamento das propriedades ACID

Armazenamento Distribuído de Dados



Bancos de Dados NoSQL

Algumas Características

- ✓ Foco no armazenamento de grandes volumes de dados
- ✓ Alto desempenho
- ✓ Alta escalabilidade horizontal
- ✓ Replicação dos dados
- ✓ SQL não é a linguagem padrão
- ✓ Usualmente classificados pelo modelo de dados
 - ❖ Chave-valor, documentos, famílias de colunas e grafos

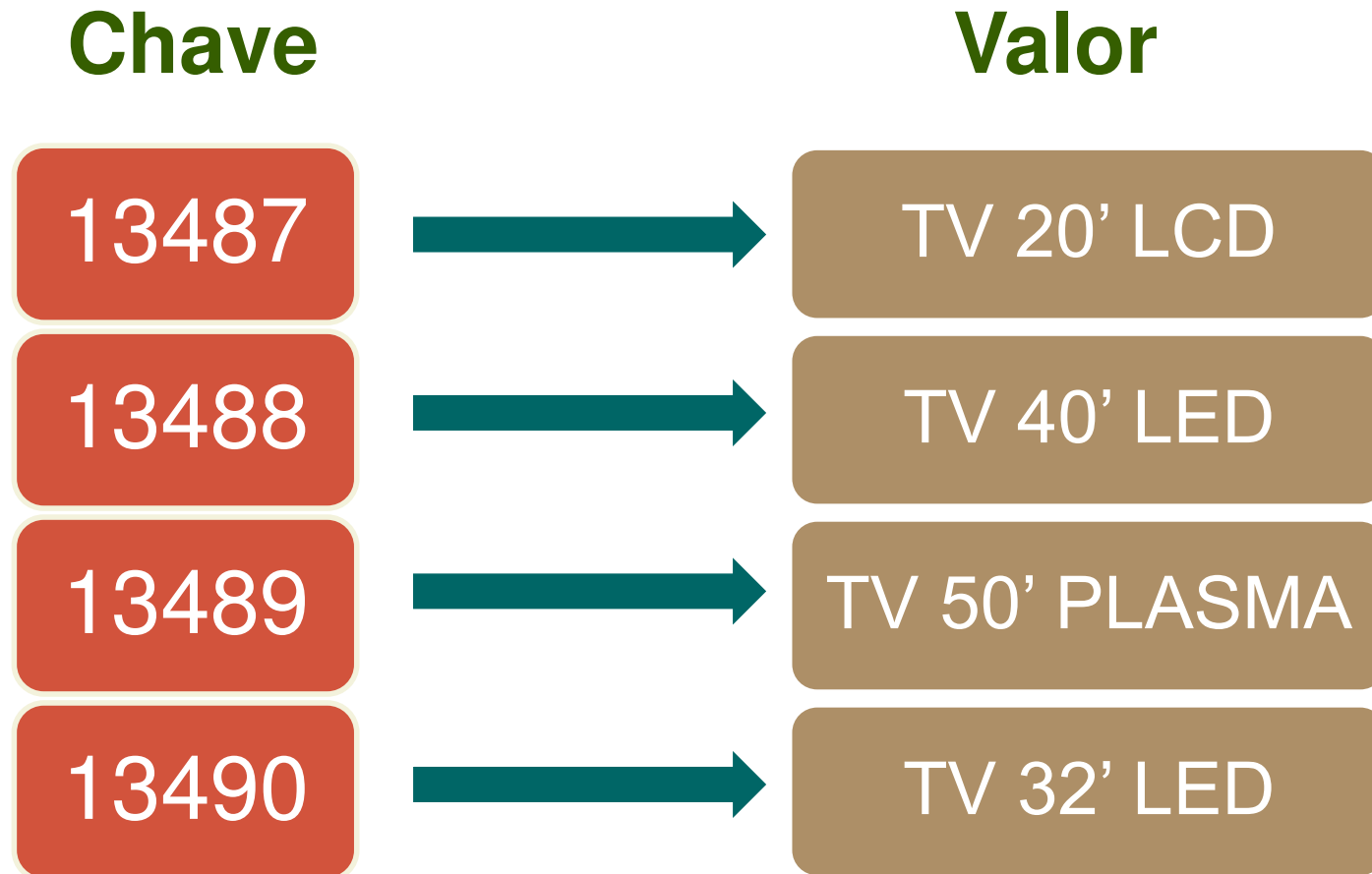
Bancos de Dados NoSQL

- **Chave-valor**

- Modelo simples
- Coleção de pares chave-valor
- Acesso aos dados pela chaves
- Pesquisa rápida
- Aplicações típicas
 - Gerenciamento de sessões



Bancos de Dados NoSQL



Bancos de Dados NoSQL

- **Documentos**

- Cada documento é uma coleção de pares chave-valor
- Formato JSON/XML
- Permite consultas mais elaboradas
 - Filtros por atributos
- Aplicações típicas
 - Aplicações Web



Bancos de Dados NoSQL

```
{  
  "Id": 13487,  
  "Produto": "TV",  
  "Tamanho": "20",  
  "Tela": "LCD"  
}
```

```
{  
  "Id": 13488,  
  "Produto": "TV",  
  "Tamanho": "40",  
  "Tela": "LED"  
}
```

```
{  
  "Id": 13489,  
  "Produto": "TV",  
  "Tamanho": "50",  
  "Tela": "PLASMA"  
}
```

```
{  
  "Id": 13490,  
  "Produto": "TV",  
  "Tamanho": "32",  
  "Tela": "LED"  
}
```

Bancos de Dados NoSQL

• Famílias de Colunas

- Eficiente ao armazenar grandes quantidades
- Valores identificados por uma tripla
 - Linha, coluna e timestamp
- Colunas são armazenadas de forma contígua
- Não implementa junção
- Aplicações típicas
 - Gerenciamento de grandes volumes de dados



Bancos de Dados NoSQL

Famílias de Colunas

Principais

Adicionais

<i>Rowkey</i>	Produto	Tamanho	Tela	Tipo
13487	TV	20'	LCD	
13488	TV	40'	LED	
13489	TV	50'	PLASMA	SMART
13490	TV	32'	LED	SMART

Bancos de Dados NoSQL

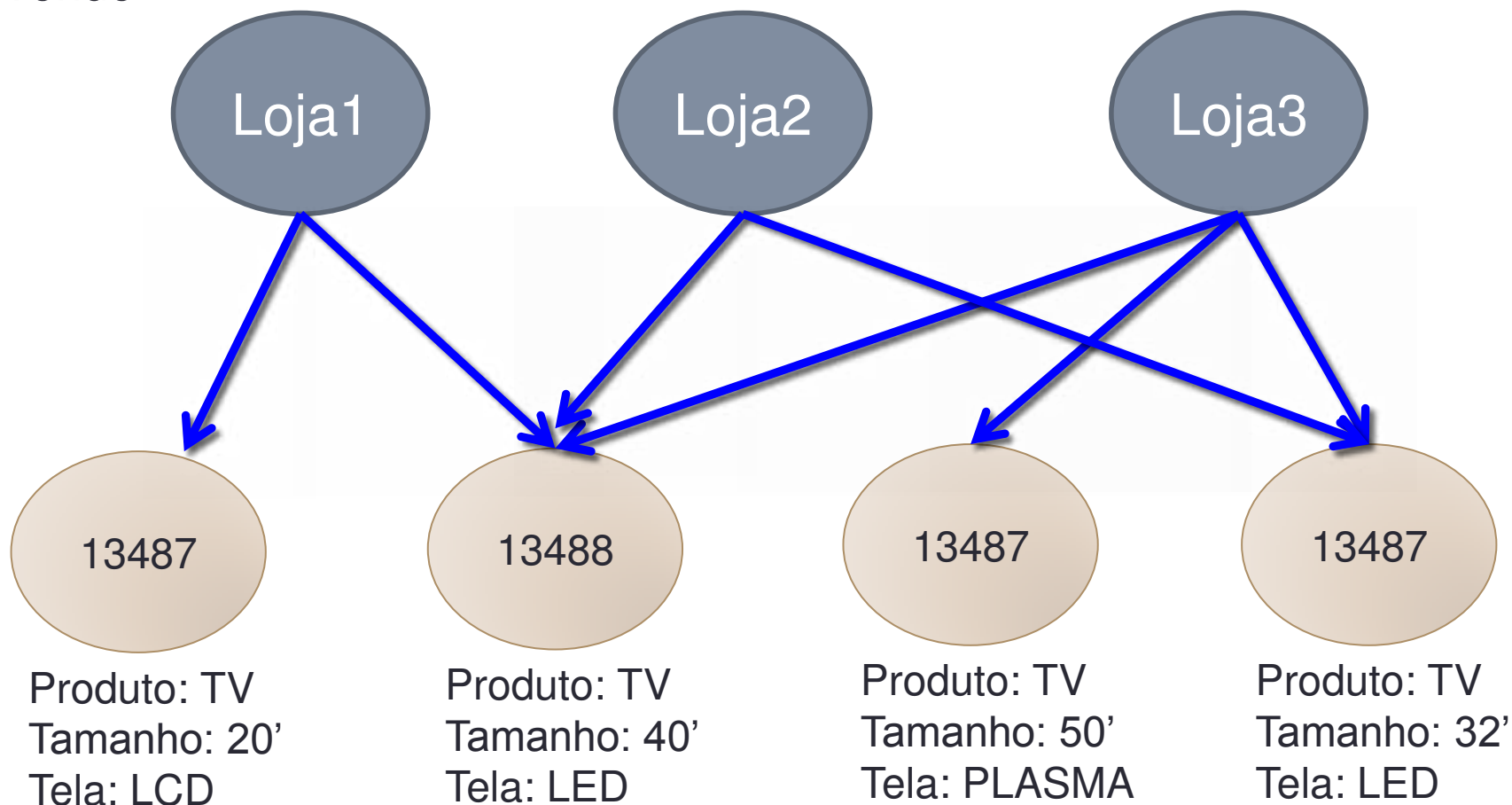
- **Grafos**

- Baseados na teoria dos grafos
- Dados são modelados por nós e arestas
- Nós e arestas podem possuir propriedades
- Aplicações típicas
 - Redes Sociais
 - Recomendações

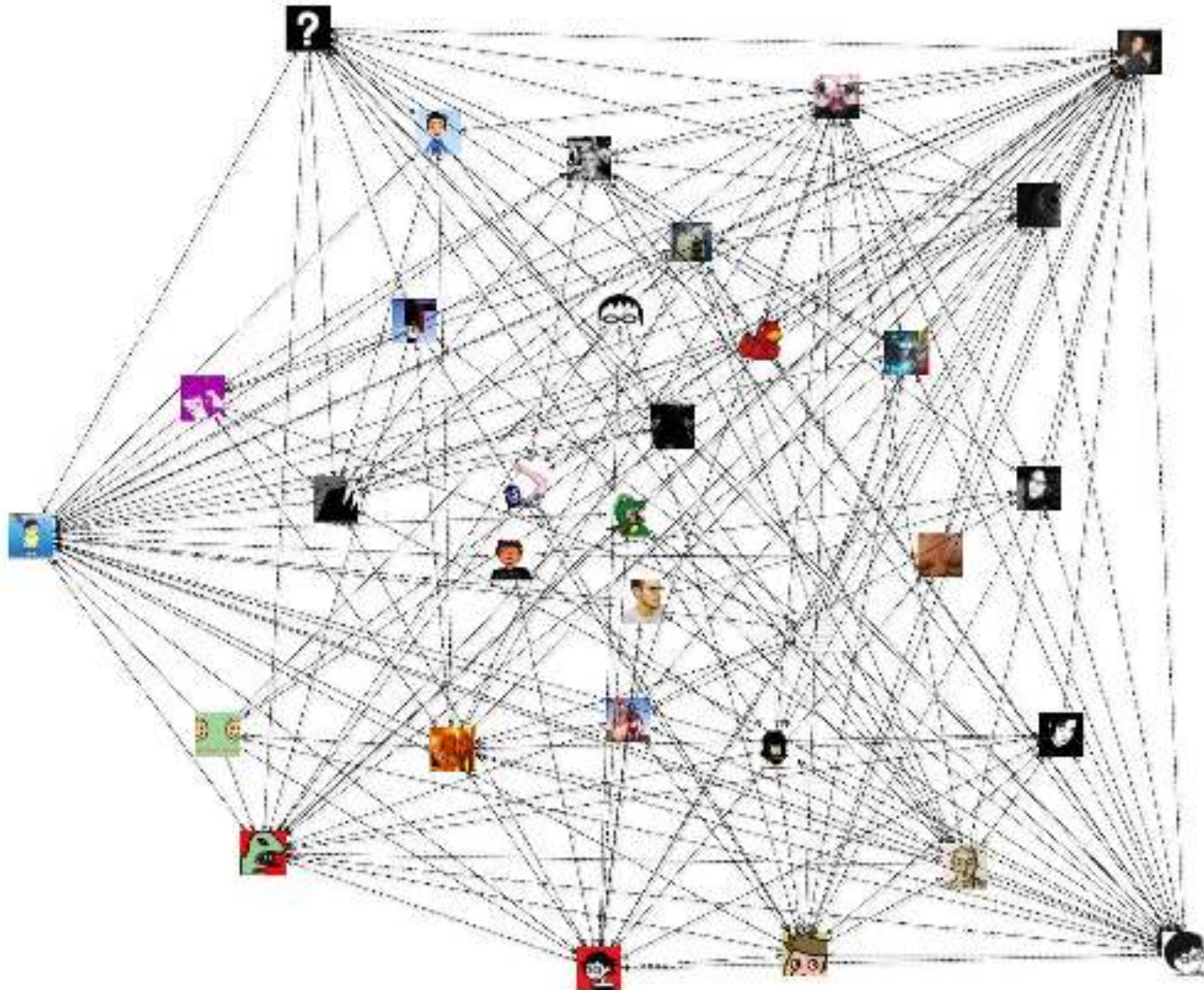


Bancos de Dados NoSQL

 Vende



Bancos de Datos NoSQL

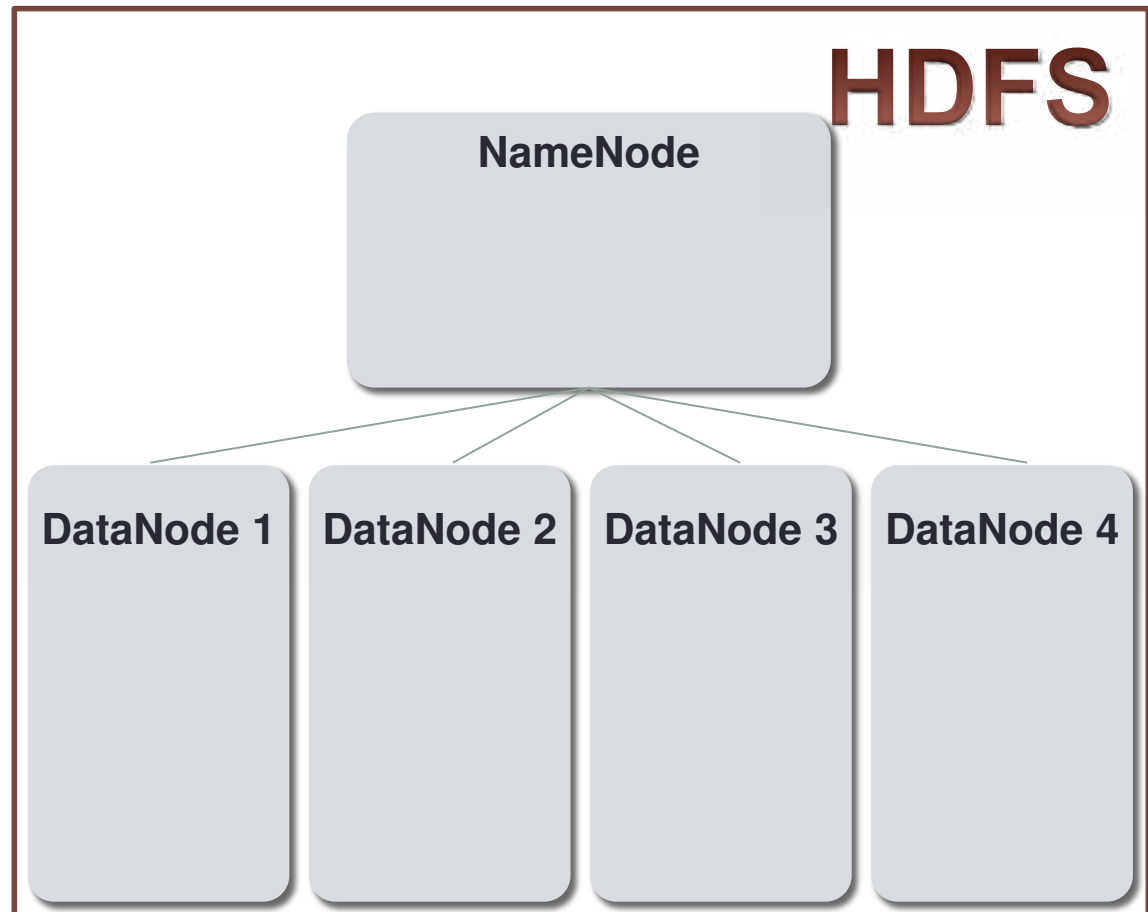


Hadoop

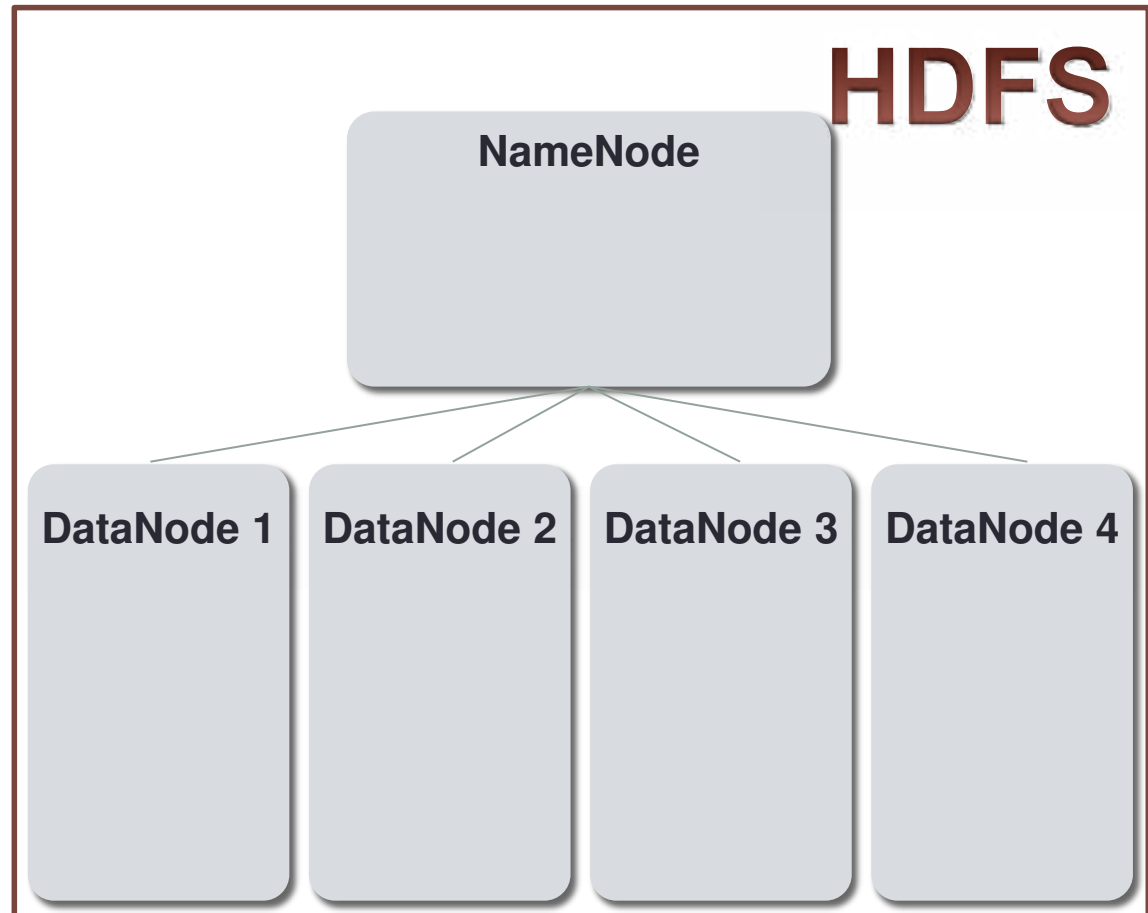
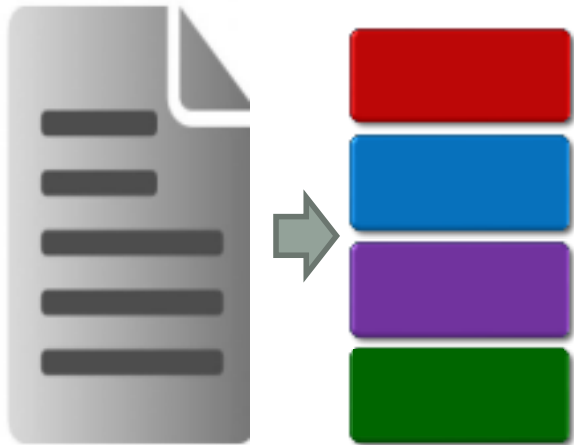
- Framework de processamento paralelo de dados em larga escala
- Altamente escalável
- Tolerante a falhas
- Disponível
- Principais componentes
 - Armazenamento
 - HDFS – Hadoop Distributed File System
 - Processamento
 - MapReduce



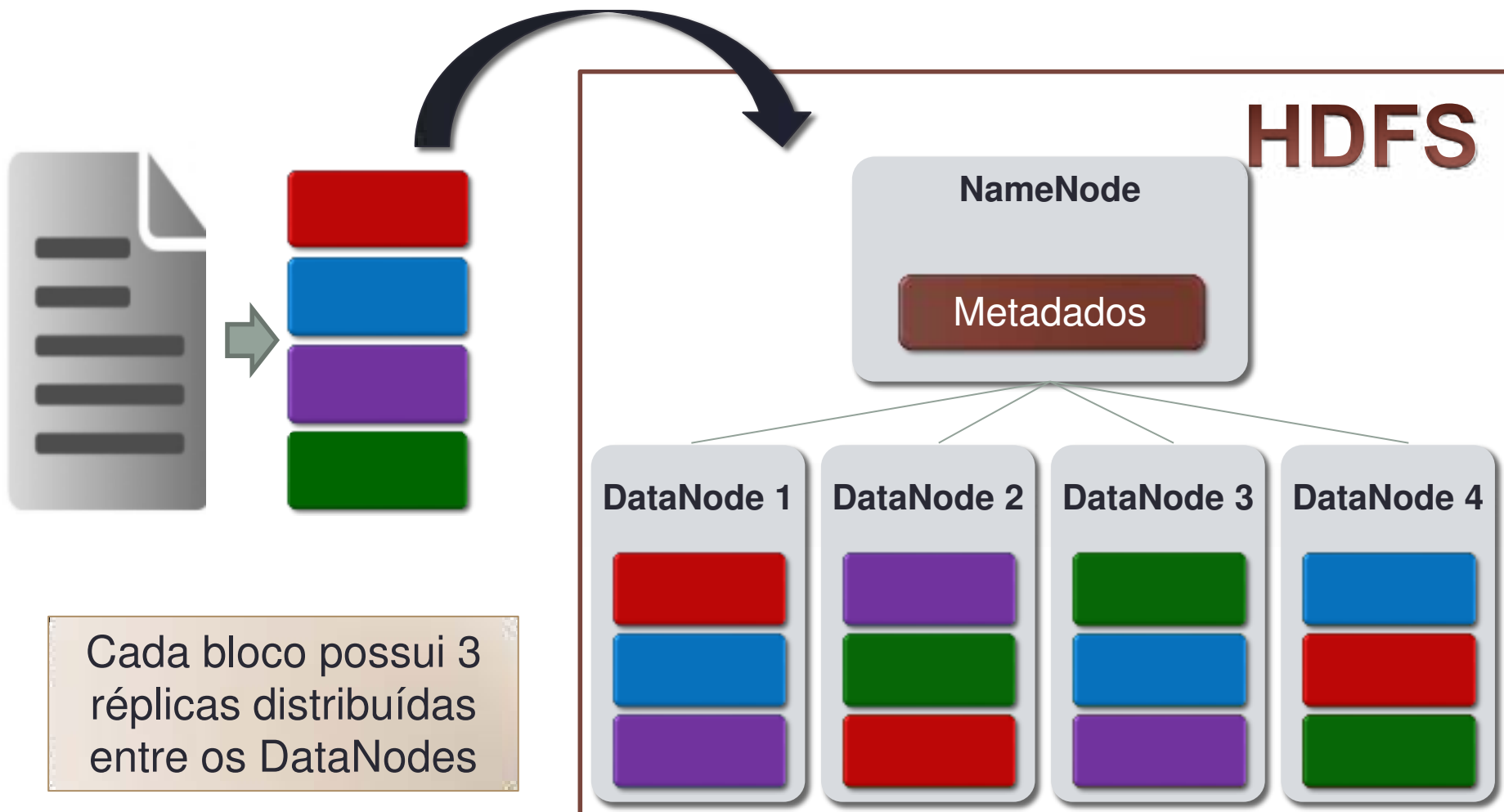
HDFS - Hadoop Distributed File System



HDFS - Hadoop Distributed File System



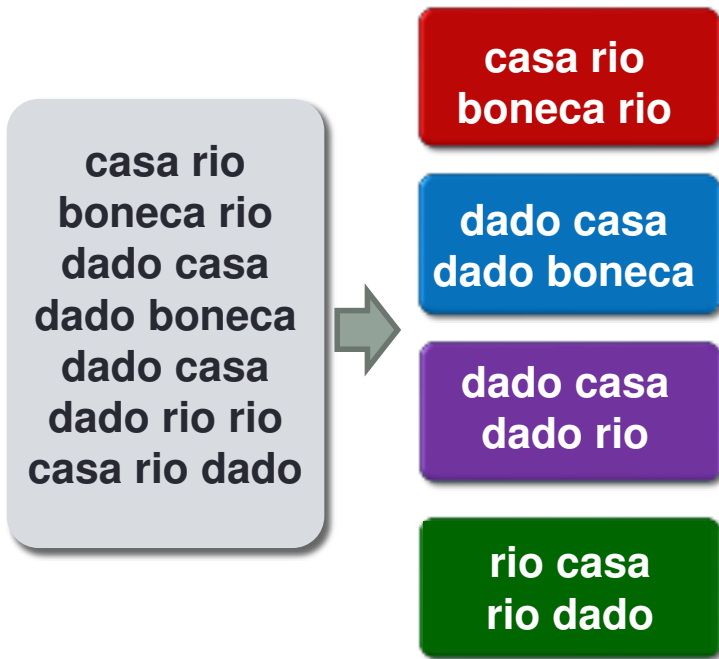
HDFS - Hadoop Distributed File System



MapReduce

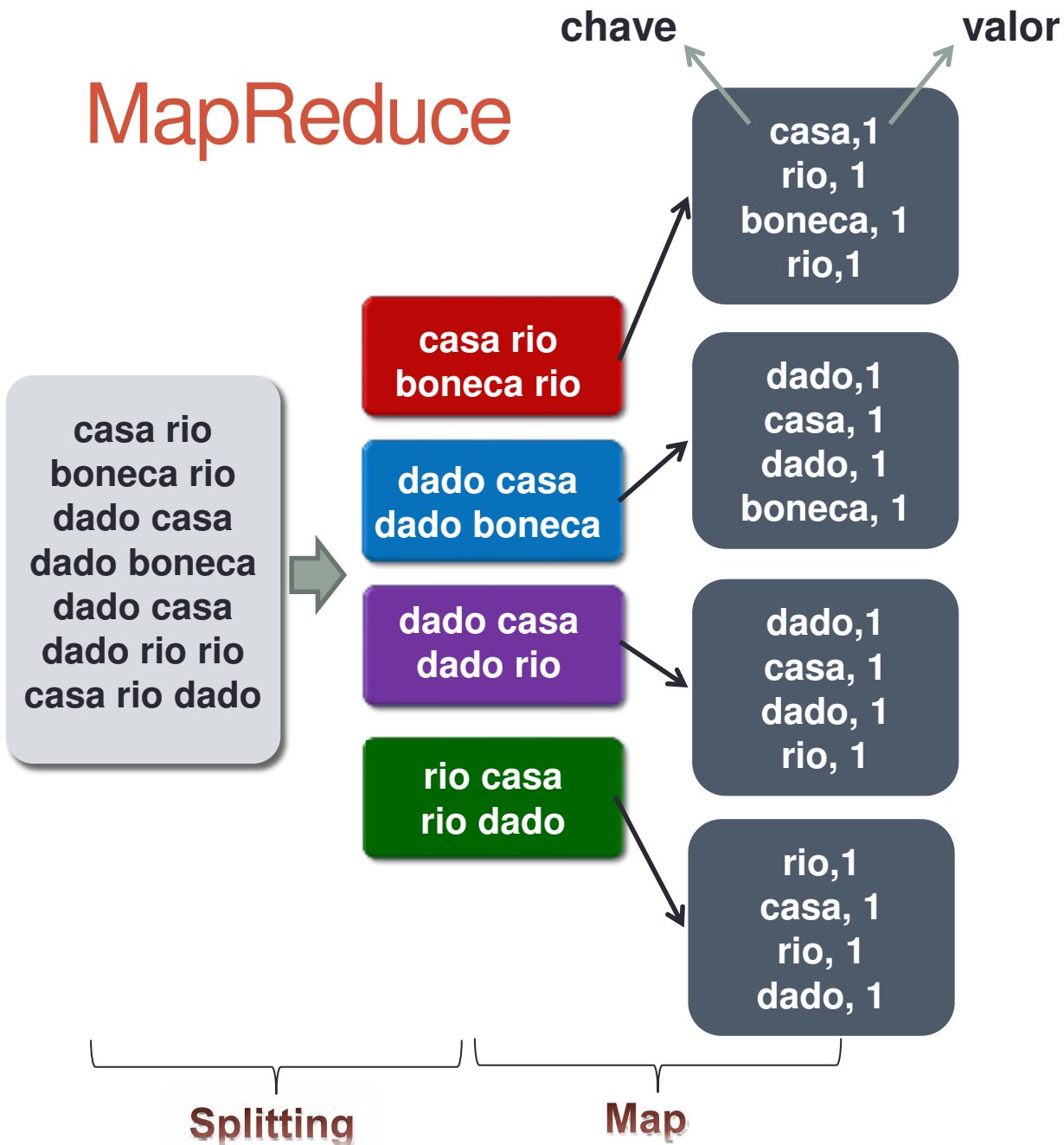
**casa rio
boneca rio
dado casa
dado boneca
dado casa
boneca rio
rio casa
boneca dado**

MapReduce

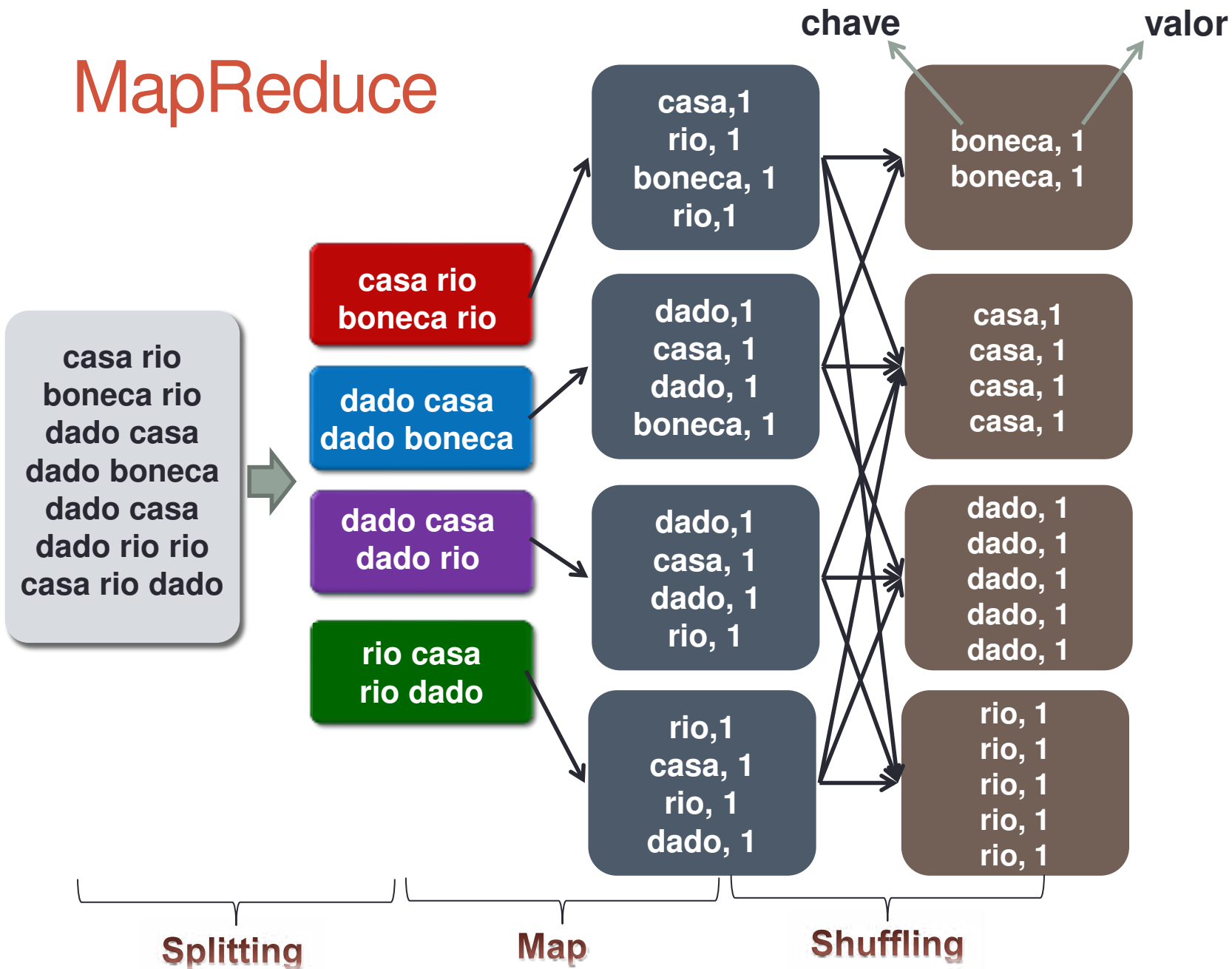


Splitting

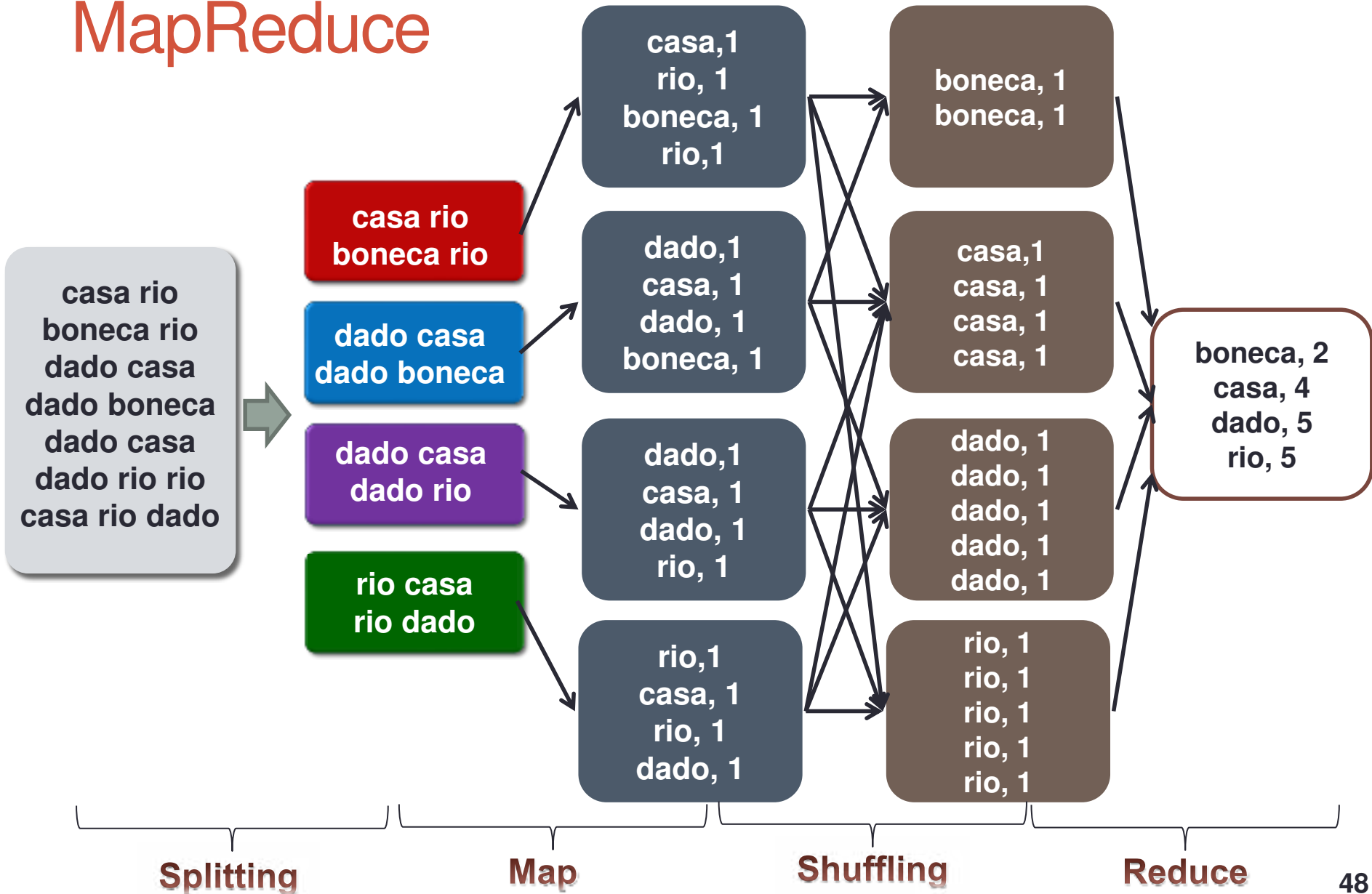
MapReduce



MapReduce



MapReduce



MapReduce

Função Map

```
public class WordCountMapper extends
    Mapper<Object, Text, Text, IntWritable> {
    private final IntWritable ONE = new IntWritable(1);
    private Text word = new Text();

    public void map(Object key, Text value, Context context)
        throws IOException, InterruptedException {
        String[] csv = value.toString().split(",");
        for (String str : csv) {
            word.set(str);
            context.write(word, ONE);
        }
    }
}
```

MapReduce

Função Reduce

```
public class WordCountReducer extends
    Reducer<Text, IntWritable, Text, IntWritable> {

    public void reduce(Text text, Iterable<IntWritable> values,
        Context context) throws IOException,
        InterruptedException {

        int sum = 0;
        for (IntWritable value : values) {
            sum += value.get();
        }
        context.write(text, new IntWritable(sum));
    }
}
```

Evolução do MapReduce



Apache Pig

- PigLatin - Linguagem procedural de alto nível
- Scripts são traduzidos para sequências de programas MapReduce

- HiveQL – Linguagem similar ao SQL para acessar os dados

Apache Hive



Oportunidades

Data Scientist

- Alguns conhecimentos necessários
 - Saber **programação**
 - Ser capaz de criar **modelos estatísticos**
 - Compreender as **diferentes plataformas de Big Data**
- Usualmente esse profissional é formado em Estatística, Matemática, Física ou Ciências da Computação



Oportunidades



Job description

Descrição da vaga

Imagine que você tenha que analisar e identificar comportamentos e tendências em uma base de mais de 60 milhões de usuários únicos, relacionando dados como idade, sexo, classe social, estado civil, geo localização e interesses de consumo classificados em mais de 180 tipos em mais de 30GB de logs por dia!

Como gerar um modelo matemático que represente esses grupos? Como criar um sistema de recomendação que aumente a rentabilidade e o aproveitamento de inventário?

Para isso estamos buscando um Data Scientist que ficará responsável por organizar a arquitetura de Big Data, analisar os dados, criar Algoritmos de Machine Learning e modelos Matemáticos que ajudem na otimização dessa plataforma.



Desired Skills and Experience

- Python, C++, R
- MySQL, MongoDB, Hadoop, ElasticSearch, Redis
- Linux, GitHub, AWS
- Técnicas de Information Retrieval e Estatística (Bayes, Markov Chain, etc)
- Sistemas de Recomendação como Recomendação Recíproca, Multi-Armed Bandit, entre outros
- Machine Learning (SVM, Deep Learning, etc)
- Inglês Fluente

Oportunidades

Data Scientist - Jersey City, NJ or Boston

Analytic Recruiting • Jersey City, NJ • 9/10/2014

★ Save Job ✉ Email 🖨 Print ⚠ Report

JOB DESCRIPTION

Data Scientist sought by major financial services firm in their Jersey City, NJ or Boston area offices. Role will provide statistical analysis, forecasting, predictive modeling, simulation, and optimization. This role will leverage Big Data mining and analysis strategies to understand customer needs, provide business insights, improve targeting, maximize return on investment, and optimize marketing efforts.

Refer to Job# 20958-CB email MS Word attached resume to Orly Miller. [[Click Here to Email Your Resumé](#)] or register online at www.analyticrecruiting.com choosing Orly Miller as your contact recruiter.

Oportunidades

Data Scientist

Vaco Technology • Santa Clara, CA • 9/9/2014

★ Save Job

✉ Email

🖨 Print

⚠ Report

JOB DESCRIPTION

Vaco has been engaged by a famous technology leader to identify a savvy **Data Scientist** for a newly formed "Data Science Team" that is building and creating the Big Data initiatives for the entire enterprise.

Join a highly visible team that is responsible for critical, high profile projects such as the initial "data cleansing" project. The Data Science Team is creating a "Data Lake" and wants to ensure that PSI (personal sensitive information including credit card, SS# details, etc) gets extracted before making its way to the lake.

Required Experience for the Data Scientist:

- 10+ years development-specific experience ("heads-down programmer")
- Expert level technical foundation including more than one advanced structured programming environment such as APEX, Force.com, Python, .Net, or Java

Preferred Experience for the Data Scientist:

- Hadoop Development

Oportunidades

Data Scientist - Big Data - Mining Massive Datasets and hacking

Laguna Source • San Francisco, CA • 8/20/2014

★ Save Job ✉ Email 🖨 Print ⚠ Report

JOB DESCRIPTION

Are you a Hacker-like Data Scientist with creative skills to gather, mine and analyze inconsistent, massive data sources? If so, read on...

*** Relocation to San Francisco is available for the winning candidate***

We are an EXTREMELY well funded company that is in the massive, multi-player on-line gaming space. Our players number in the MILLIONS. Our investors are a who's who of Media and Tech.

Top Reasons to Work with Us

- Work as a DATA GURU for a team working on bleeding edge ideas
- Build out tubing that enables the transport/processing/normalization of MASSIVE data
- Build and maintain complex statistical models that learn from and scale to millions of users

What You Will Be Doing

- Mine with multiple imperfect, mixed, varied, and inconsistent data sources (Technical expertise needed in one or more of the following languages/tools to wrangle and understand data: Python, SQL, Hive, R, Matlab, Spotfire, Tableau)
- Automate the import of data from a variety of sources (ad providers, for example) into a singular data source.

Referências

<http://www.sciencedaily.com/releases/2013/05/130522085217.htm>

http://www-3.unipv.it/ingegneria/copisteria_virtuale/motta/doss/13-PV-DoES-18-BigData-v3.pdf

<http://pig.apache.org/>

<http://hadoop.apache.org/>

<https://hive.apache.org/>

<http://azure.microsoft.com/en-us/pricing/details/storage/>

<http://thoughtsoncloud.com/2014/02/top-7-most-common-uses-of-cloud-computing/>

<http://pt.slideshare.net/laodias/os-cinco-vs-do-big-data>

<http://www.microsoft.com/en-us/news/presskits/cloud/docs/the-economics-of-the-cloud.pdf>

<http://www.cloudproviderusa.com/the-difference-between-public-cloud-private-cloud/>

<http://observatorio.inweb.org.br/dengueapp?next=/dengueapp/relatorio>

<https://www.youtube.com/watch?v=jyx8iP5tfCI&feature=youtu.be>

Obrigado 😊